# Cybersecurity in the Era of Quantum Computing and Advanced AI: Emerging Threats and Future Directions

**Hisham Alnabulsiyyah and Meteabah Aldawsari**
College of Computer and Information Technology, Shaqra University, Saudi Arabia

abo3tb0092@gmail.com

**Abstract**: Quantum computing and advanced artificial intelligence are beginning to disturb several assumptions on which current cybersecurity practice depends. This paper reviews two related areas of risk: the possibility that cryptographically relevant quantum computers could weaken widely deployed public-key encryption, and the ways in which generative AI and large language models are already changing cyber operations. The analysis finds that quantum risk should not be treated as a distant event only, since sensitive encrypted data can be collected now and decrypted later if suitable quantum capability becomes available. It also finds that advanced AI is already affecting the scale, speed, and credibility of social engineering, reconnaissance, vulnerability analysis, and misinformation. The paper discusses the current state of these threats, outlines plausible future scenarios, and identifies practical defensive priorities, including post-quantum cryptography migration, AI system hardening, red teaming, content authentication, workforce training, and governance. The central argument is that quantum and AI security should be handled as near-term resilience issues, even where their most severe effects may appear later.

**Keywords**: Cybersecurity, Quantum computing, Post-quantum cryptography, Artificial intelligence, Large language models, Deepfakes

## 1. Introduction

Cybersecurity has long relied on three assumptions: that certain computations are impractical for attackers, that human users can recognize some forms of deception, and that defenders can observe and respond to compromise before damage becomes systemic. Quantum computing and advanced AI put pressure on all three. Quantum computing threatens the mathematical basis of many public-key systems, while generative AI gives attackers a practical way to produce convincing language, code, audio, and images at scale. These developments do not replace existing cyber risks, instead, they make familiar attacks faster, cheaper, and in some cases harder to recognize.

This paper examines how these technologies reshape the cyber threat landscape and which responses should be prioritized. It focuses on two questions. First, how could quantum computing affect encryption, signatures, and digital trust? Second, how could advanced AI and large language models introduce new vulnerabilities or strengthen existing attacker capabilities? The discussion draws on recent academic, government, and industry sources, with emphasis on realistic security implications rather than speculative forecasts. The paper does not attempt to predict an exact date for a cryptographically relevant quantum computer. That date remains uncertain. The more useful question is whether the direction of risk already justifies defensive action.

The argument advanced here is that the security community is already in a transition period. Quantum computers are not currently breaking strong, real-world public-key encryption, but migration to quantum-resistant systems will be slow and operationally difficult. Current AI systems are also not fully autonomous attackers, but they already improve phishing, impersonation, malware support, and vulnerability analysis. Waiting for either threat to mature before preparing would leave organizations with limited time to respond. A more defensible approach is to build cryptographic agility, secure AI development practices, and governance controls while the risks are still manageable.

## 2. Quantum Computing and the Cryptographic Risk

The most widely discussed cybersecurity implication of quantum computing is its effect on public-key cryptography. Systems such as RSA, Diffie-Hellman, and elliptic-curve cryptography depend on mathematical problems that are difficult for classical computers, including integer factorization and discrete logarithms. A sufficiently powerful quantum computer running Shor's algorithm could solve these problems efficiently, making many current public-key schemes unsafe [1], [2]. The effect would reach far beyond encrypted messaging. Secure web browsing, virtual private networks, software updates, digital signatures, identity systems, and parts of financial infrastructure all rely on these mechanisms.

The more immediate concern is not only future decryption. The strategy often described as harvest now, decrypt later changes the timeline. An adversary can collect encrypted traffic or stored encrypted records

today and preserve them until quantum decryption becomes practical. Data with long confidentiality value, such as government communications, medical records, intellectual property, legal material, and financial information, may therefore be exposed before the technical ability to decrypt it exists [3]. This is why organizations that protect long-lived sensitive data cannot wait for a public quantum breakthrough before planning their transition.

Current quantum hardware still faces major limits, including noise, error rates, and the difficulty of scaling fault-tolerant systems. Publicly known systems do not yet have the stable, error-corrected capability required to break large commercial cryptographic deployments. However, progress in algorithms, error correction, and hardware engineering has reduced confidence in the assumption that the risk is safely remote. NIST and related agencies have warned that preparation should begin before the threat becomes operational [1], [3]. The practical challenge is that cryptographic migration is not a single software patch. It requires finding where cryptography is used, replacing vulnerable algorithms, testing interoperability, updating certificates and protocols, managing legacy devices, and coordinating with vendors.

Post-quantum cryptography is the main defensive pathway. These algorithms are designed to resist both classical and quantum attacks by relying on mathematical assumptions such as structured lattices, hash-based signatures, and code-based problems. NIST's standardization process has produced finalized post-quantum encryption and signature standards, which gives organizations a clearer basis for planning [1]. The transition should be managed through cryptographic agility: systems should be designed so that algorithms can be replaced without redesigning the full architecture. That matters because post-quantum algorithms are newer than the schemes they will replace, and implementation guidance will continue to mature. The most sensitive and long-lived assets should be addressed first.

Quantum risk also affects trust models beyond confidentiality. Digital signatures verify software updates, contracts, identity credentials, and blockchain transactions. If widely used signature schemes become vulnerable, attackers could impersonate trusted parties or distribute malicious updates that appear legitimate. Symmetric cryptography is less exposed, but it is not unaffected, Grover's algorithm reduces the effective security strength of symmetric keys, which means key sizes and hash functions should also be reviewed. These issues do not make current systems obsolete overnight, but they do require a systematic inventory and staged upgrade plan.

## 3. Advanced AI and Large Language Model Security Risks

Compared with quantum computing, advanced AI is already operationally relevant because usable tools are widely available now. Generative models can draft convincing emails, summarize stolen documents, translate scams, imitate a professional tone, generate code, and support research on targets. Large language models also create new attack surfaces when they are embedded into customer service tools, coding assistants, search systems, document analysis tools, and autonomous agents. The main security issue is not that every model output is dangerous, but that models can be connected to sensitive data and actions without the same predictable boundaries as conventional software.

Prompt injection is one of the clearest examples. In a prompt injection attack, malicious instructions are placed inside user input, a document, a web page, or another data source that the AI system processes. The model may then follow those instructions rather than the developer's intended policy. In practice, an attacker could try to cause an AI assistant to disclose sensitive content, ignore access rules, send data externally, or take an unsafe action through a connected tool [4], [5]. The risk grows when an LLM has access to email, files, databases, or internal applications, because the model may not reliably separate trusted system instructions from untrusted content that merely looks like an instruction.

Other AI-specific risks include jailbreaking, adversarial examples, data poisoning, and model backdoors. Jailbreaking attempts to bypass restrictions and obtain outputs that the system is meant to block. Adversarial examples manipulate inputs so that a model behaves incorrectly. Data poisoning targets training or fine-tuning data, while backdoors remain hidden until a specific trigger appears. These problems are difficult to manage because modern models are complex, probabilistic, and not fully interpretable. A system can perform well in ordinary testing and still fail when placed under adversarial pressure.

AI also changes the economics of conventional cybercrime. Phishing is the most visible case. Older phishing campaigns often contained weak grammar, generic wording, or obvious inconsistencies. Generative AI allows attackers to produce personalized messages that sound professional and can be adapted quickly after failure. The FBI has warned that criminals are using AI to increase the speed, scale, and believability of fraud and social

engineering [6]. The same pattern applies beyond email, including text messages, voice calls, fake job offers, business email compromise, and customer support fraud.

Deepfakes extend the same problem into voice and video. AI-generated media can imitate executives, relatives, public figures, or colleagues. In a corporate setting, a convincing fake voice call can support payment fraud or credential theft. In a public setting, fake video or audio can spread misinformation, damage reputations, or create short-term panic before verification catches up. Detection tools are useful, but they are not enough on their own, since attackers can improve generation quality and exploit situations where people are under time pressure.

AI may also accelerate technical attacks. Models can assist with code analysis, vulnerability research, exploit adaptation, and the interpretation of technical documentation. They are unlikely to replace expert attackers in the near term, but they can lower the skill required for some tasks and increase productivity for experienced teams. A threat actor can use AI to generate reconnaissance scripts, analyze error messages, rewrite simple malware, or summarize the configuration of a target technology. The expected defensive challenge is therefore greater volume, faster iteration, and more convincing social engineering rather than a single new form of attack.

## 4. Combined Threat Landscape and Future Scenarios

Quantum and AI risks are often discussed separately, but operationally they are part of a broader shift. AI gives attackers automation and persuasion advantages now, while quantum computing threatens the cryptographic layer that supports digital trust over the longer term. The combined effect is a security environment in which attackers may become better at gaining access today and, in the future, better able to exploit information that was previously protected by older cryptography.

One plausible scenario begins with data harvesting. An attacker collects encrypted records from a high-value organization and stores them. AI-assisted tools are then used to identify employees, suppliers, and business processes, making credential theft or supplier impersonation more credible. Once inside a network, the attacker searches for stored encrypted databases, certificate material, and long-term secrets. If quantum decryption later becomes practical, data taken years earlier may become readable. This scenario links a present compromise to a future cryptographic failure and shows why the two risks should be planned together.

A second scenario involves AI-enabled exploitation of AI systems themselves. Organizations are increasingly deploying chatbots and assistants into workflows that can access documents, calendars, customer records, or code repositories. If these systems lack strong boundaries, indirect prompt injection can become a route from an untrusted document to a sensitive action. A malicious document, for example, may instruct an assistant to summarize confidential files or leak selected information through an external request. The model does not need to be malicious, it only needs to become the mechanism through which the attack is carried out.

Critical infrastructure raises the stakes further. AI systems may support monitoring, optimization, and decision-making in energy, healthcare, transport, and finance. At the same time, long-lived infrastructure communications may require quantum-safe protection. The risks are not limited to confidentiality. Integrity failures can be equally serious: a manipulated AI recommendation, a forged signature, or a compromised update process could cause operators to trust false information.

It is also important not to frame either technology only as a threat. AI can improve anomaly detection, malware triage, incident response, and vulnerability management. Quantum technologies may eventually support stronger forms of key distribution or random number generation in specialized settings. These benefits will only be useful if deployment is disciplined. Without clear controls, validation, and accountability, the same capabilities can increase complexity and enlarge the attack surface.

## 5. Mitigation Strategies

For quantum risk, the practical starting point is an inventory. Organizations need to know where public-key algorithms are used, which systems protect long-lived data, and which vendors control cryptographic components. The inventory should include certificates, VPNs, web services, email encryption, identity systems, code-signing processes, embedded devices, and backup systems. After that, assets should be classified by urgency. Systems protecting sensitive data for many years should move earlier in the migration plan than systems with short-lived confidentiality needs.

Migration to post-quantum cryptography should be staged and tested. Hybrid approaches may be useful during the transition because they combine classical and post-quantum mechanisms while the newer schemes mature in operational use. Procurement requirements should also be updated so vendors demonstrate post-quantum support and cryptographic agility. This is especially important for devices and platforms with long service lives, since they may still be deployed when quantum threats become more practical.

For AI security, LLM applications should be treated as software systems with unusual failure modes, not as simple user interfaces. Access control should be enforced outside the model. The model should not be the only authority deciding whether a user may view, send, or modify information. Sensitive actions should rely on deterministic policy checks, logging, and human confirmation where appropriate. External documents, websites, and emails should be treated as untrusted inputs even when processed by an internal AI assistant.

Prompt injection defenses need layers. Separating system instructions from user content is useful, but it is not sufficient by itself. Applications should limit which tools the model can call, restrict data access to the minimum necessary, sanitize or transform untrusted inputs, and monitor outputs for suspicious behavior. Red teaming should be part of the deployment lifecycle. AI systems should be tested with adversarial prompts, malicious documents, indirect prompt injection examples, jailbreak attempts, and realistic abuse cases before and after release.

Social engineering controls also need to be updated. Employees should be trained on AI-generated phishing and deepfakes, but training cannot be the only control. High-risk actions such as payments, credential resets, sensitive data exports, or urgent executive requests should require independent verification through trusted channels. Phishing-resistant multi-factor authentication, domain authentication, and transaction approval workflows remain necessary because awareness alone is unreliable when people are under pressure.

Content authenticity is another defensive layer. Watermarking, provenance metadata, and cryptographic signing of legitimate media can help users and institutions verify whether content came from a trusted source. These methods will not stop all deception, but they can reduce uncertainty in high-value communications. Incident response plans should also include deepfake procedures, including rapid verification, public communication, and escalation paths.

Governance connects these technical controls to responsibility. AI systems used in security-sensitive contexts should have documented owners, risk assessments, logging, model update procedures, and clear rules for human oversight. Quantum readiness should be included in enterprise risk management rather than left as a research topic for specialists. Organizations that prepare early will have more control over cost, compatibility, and timing than those that wait for a crisis.

## 6. Future Research Directions

A first research priority is practical migration engineering for post-quantum cryptography. Standards are only the starting point. Researchers and practitioners need evidence about performance, interoperability, certificate sizes, implementation mistakes, and secure deployment patterns in real systems. Constrained devices, industrial systems, and legacy infrastructure deserve particular attention because many cannot be upgraded quickly or cheaply.

A second priority is AI system evaluation. Current benchmarks do not fully capture security behavior in real deployments, especially when models are connected to tools and private data. Future evaluations should test indirect prompt injection, multi-step attacks, tool misuse, data leakage, model behavior under conflicting instructions, and resistance to adaptive jailbreaks. The goal is not to prove that a model is perfectly safe, which is unrealistic, but to measure risk clearly enough to support deployment decisions.

A third priority is interpretability and verification. Security teams need better ways to understand why a model takes a certain action and whether hidden instructions or poisoned data influenced it. This is especially important for autonomous agents that plan across multiple steps. If an AI system can read files, call APIs, and make decisions, defenders need audit trails and policy enforcement that are stronger than ordinary chat logs.

A fourth priority is human-centered security. AI-generated deception targets trust, urgency, and authority. Research should examine which verification methods work under realistic pressure and which training designs actually change behavior. Technical systems should support people rather than assume that every individual will detect every fake. For example, payment systems can require independent confirmation by design instead of relying on an employee to notice a suspicious voice call.

Finally, cross-disciplinary education is needed. Cybersecurity professionals will need enough knowledge of AI and quantum computing to ask the right questions, while AI and quantum specialists will need enough security knowledge to understand misuse and deployment risk. Universities, professional programs, and industry training should reflect this shift. The future security workforce will need both technical specialization and the ability to work across disciplines.

## 7. Current State and Future Outlook

The current quantum threat is best described as a preparation problem before the main impact. Today's quantum computers are not able to break large public-key systems used across ordinary digital infrastructure. The risk is therefore not an immediate mass decryption event. The risk is that migration will be long, uneven, and dependent on many vendors and protocols. Organizations that begin with inventory and staged testing now can move deliberately. Organizations that wait for a major breakthrough may face rushed migration, compatibility failures, and continued exposure of archived data.

AI risk is at a different stage. AI-enabled attacks are already visible because widely available tools can produce text, images, voice, and code. The most realistic near-term expectation is not a fully autonomous AI hacker, but a steady increase in the efficiency of human-led attacks. This matters because cybersecurity teams are already overloaded by alert volume, patching demands, and social engineering. AI adds pressure to the same weak points.

Over the next five to ten years, the defensive gap will likely depend less on whether organizations have heard about these risks and more on whether they convert awareness into engineering change. Quantum readiness will require procurement rules, protocol updates, and support from cloud providers, software vendors, and device manufacturers. AI readiness will require secure design patterns for AI applications, stronger identity verification, and incident response for synthetic media. Organizations that treat these areas as ordinary cyber hygiene will reduce disruption. Those that treat them as distant research topics may find that their controls age faster than expected.

## 8. Conclusion

Quantum computing and advanced AI are reshaping cybersecurity in different but connected ways. Quantum computing threatens public-key cryptography and makes post-quantum migration necessary, especially for data that must remain confidential for many years. Advanced AI is already increasing the quality and scale of cyber threats, particularly in phishing, impersonation, vulnerability support, and exploitation of AI-enabled systems. Both technologies create risks that traditional controls alone cannot fully address.

The most important issue is timing. Waiting for quantum computers to break encryption or for AI attacks to become fully automated would be a strategic error. Post-quantum migration requires years of inventory, testing, vendor coordination, and operational change. AI security requires stronger access boundaries, red teaming, secure development practices, and governance. Defenders still have a window to prepare, but that window will narrow as the technologies mature.

A practical cybersecurity strategy for the next decade should therefore be both quantum-aware and AI-aware. It should protect long-lived data, build cryptographic agility, secure AI systems as part of the software supply chain, verify high-risk communications, and train professionals who can work across these domains. The aim is not to resist technological progress, but to keep digital infrastructure trustworthy as the underlying technology changes.

**Ethics Statement:** This paper is a literature-based cybersecurity study and does not involve human participants, animal subjects, clinical data, personal data collection, experiments on live systems, unauthorized access, malware deployment or active testing against third-party infrastructure. The discussion of threats is presented at a high level for academic and defensive purposes. The paper avoids operational instructions that would enable cyber abuse and is intended to support responsible security planning, risk awareness and future research.

**AI Use Statement**: AI-assisted editing tools were used in a limited way during revision to support grammar, clarity, formatting, and consistency checks. The authors reviewed all arguments, verified the cited sources, and accept full responsibility for the accuracy, originality, and integrity of the final manuscript. No AI tool was used as an independent source of authority for the technical claims in this paper.

## References

Cybersecurity and Infrastructure Security Agency (CISA), National Security Agency (NSA), and National Institute of Standards and Technology (NIST), Quantum-Readiness: Migration to Post-Quantum Cryptography, Aug. 2023. [Online]. Available: https://www.cisa.gov/sites/default/files/2023-08/Quantum%20Readiness_Final_CLEAR_508c%20%283%29.pdf. Accessed: May 8, 2026.

E. Tabassi, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, National Institute of Standards and Technology, Jan. 2023, doi: 10.6028/NIST.AI.100-1. [Online]. Available: https://doi.org/10.6028/NIST.AI.100-1.

European Union Agency for Cybersecurity (ENISA), ENISA Threat Landscape 2024, Sep. 19, 2024. [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024. Accessed: May 8, 2026.

Federal Bureau of Investigation, "FBI Warns of Increasing Threat of Cyber Criminals Utilizing Artificial Intelligence," May 8, 2024. [Online]. Available: https://www.fbi.gov/contact-us/field-offices/sanfrancisco/news/fbi-warns-of-increasing-threat-of-cyber-criminals-utilizing-artificial-intelligence. Accessed: May 8, 2026.

IBM, "How a new wave of deepfake-driven cyber crime targets businesses," IBM Think, 2024. [Online]. Available: https://www.ibm.com/think/insights/new-wave-deepfake-cybercrime. Accessed: May 8, 2026.

M. Q. Li and B. C. M. Fung, "Security concerns for Large Language Models: A survey," Journal of Information Security and Applications, vol. 95, article 104284, Dec. 2025, doi: 10.1016/j.jisa.2025.104284.

National Cyber Security Centre, The Near-Term Impact of AI on the Cyber Threat, Jan. 24, 2024. [Online]. Available: https://www.ncsc.gov.uk/report/impact-of-ai-on-cyber-threat. Accessed: May 8, 2026.

National Institute of Standards and Technology (NIST), "NIST Releases First 3 Finalized Post-Quantum Encryption Standards," Aug. 13, 2024. [Online]. Available: https://www.nist.gov/news-events/news/2024/08/nist-releases-first-3-finalized-post-quantum-encryption-standards. Accessed: May 8, 2026.

OWASP Foundation, OWASP Top 10 for Large Language Model Applications 2025, Nov. 2024. [Online]. Available: https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/. Accessed: May 8, 2026.

P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in Proceedings 35th Annual Symposium on Foundations of Computer Science, Santa Fe, NM, USA, Nov. 20–22, 1994, pp. 124–134, doi: 10.1109/SFCS.1994.365700.